Métodos Bayesianos - Lista 1

(1) Diferenças Inferência Clássica x Bayesiana:

Clássica: Parâmetros são valores fixos desconhecidos.
não são v.a. portanto não fazemos declarações sobre sua probabilidade.

Desconsidera conhecimento prévio.

Uso da verossimilhança como base.

Bayesiana: Parâmetros são v.as → têm prob. associada.

Usa "crenças" (a-priori) combinadas com a verossimilhança para obter a posterior: (distribuição associada ao parâmetro de interesse).

— // —

(2) Vantagem da abordagem Bayesiana:

Interpretação Direta do parâmetro → Temos sua distribuição a posteriori.

Exemplo:

Bayes → "Há 95% de chance de θ estar no intervalo (A, B)"

freq → "Com a repetição infinita do experimento 95% dos intervalos de confiança conteriam θ"

PS: Em Bayesiana não temos intervalo de confiança mas sim intervalo de credibilidade.

② a) a priori subjetiva:

Distribuição de probabilidade que expressa a crença/conhecimento prévio geralmente de um especialista.

b) a priori conjugada:

"Uma distribuição a priori é conjugada para uma função de verossimilhança se a posteriori resultante pertencer a mesma família da a priori."

c) Diferença entre informativa e não-informativa

Informativa: Projetada para refletir na a posteriori informações prévias sobre o parâmetro de interesse. geralmente variância pequena e concentrada em certos valores + prováveis.
Não Informativa: Tem impacto mínimo na posteriori.

d) Própria: Uma densidade $p(\theta)$ é própria se em todo domínio de θ, $\int_{\theta} p(\theta) = 1$, caso a integral seja finita $\neq 1$ pode ser normalizada para se tornar própria.

Imprópria: se $\int_{\theta} p(\theta)$ não integrar em valor finito então é imprópria.

Queremos: Temos:

(1) $P(I|G) = \dfrac{P(G|I) \cdot P(I)}{P(G)}$ $P(I) = 1/200$, $P(F) = 1/125$

$P(G|I) \to$ serem dois irmãos dado que são idênticos:

$\therefore P(G|I) = 1/2$

$P(G|F) \to$ serem irmãos | são fraternos. $P(G|F) = 1/4$

Prob. Total: Dado Nascimento de Gemeos $\to$ 2 meninos

$P(G) = P(G|I) \cdot P(I) + P(G|F) \cdot P(F) = 0.003\overline{6}$

$\therefore P(I|G) = \dfrac{1/2 \cdot 1/300}{0.003\overline{6}} \cong \underline{0.45}$ A probabilidade é de 0.45.

Exercício 5 → Bayesiana

a) Núcleo Normal $N(\mu, \sigma^2)$

O núcleo de $N(\mu, \sigma^2)$ é ~~da~~ fdp

$$f(x \mid \mu, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{A} \, e^{\underbrace{-\frac{(x-\mu)^2}{2\sigma^2}}_{B}} \to \text{NÚCLEO}$$

A → constante (ajusta altura da curva)

B → Esse é o núcleo! parte funcional que depende de x e dá a distribuição a forma de sino

b) Núcleo de Beta (α, β)

fdp da Beta:

$$f(x \mid \alpha, \beta) = \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)}}_{A} \underbrace{x^{\alpha-1}(1-x)^{\beta-1}}_{B}$$

A → constante

B → Núcleo = $x^{\alpha-1}(1-x)^{\beta-1}$

c) Núcleo Gama (α, β)

fdp:

$$f(x \mid \alpha, \beta) = \underbrace{\frac{\beta^\alpha}{\Gamma(\alpha)}}_{A} \underbrace{x^{\alpha-1} e^{-\beta x}}_{B}$$

A → constante

B → Núcleo

d) Núcleo Poisson (λ)

~~fdp~~ fmp → (é discreta, então não é fdp)

$$P(X = k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

A → núcleo

B → constante

Exercício 6) X_i, VA iid com dist Poisson (λ) tal que λ é a priori Gama (a,b) com hiper parâmetros a, b conhecidos

a) PASSO 1 Encontrar posteriori, ou seja $P(\lambda \mid \text{dados}) \propto P(\text{dados} \mid \lambda) \cdot P(\lambda) \rightarrow$ núcleo a priori

como λ é nossa variável e substituimos no núcleo de Gama $\therefore$ $P(\lambda) \propto \lambda^{a-1} e^{-b\lambda}$

$\therefore$ $P(X_i = x_i \mid \lambda) = \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!}$ pois nossos dados $(X_1, \dots, X_n)$ vem de uma Poisson (λ)

→ prob de uma única obs

PASSO 2 → multiplicar isso para todas n observações

verossimilhança → $P(\text{dados} \mid \lambda) = \displaystyle\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_1^n x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$ → constante! não depende de λ

$\therefore$ $P(\text{dados} \mid \lambda) \propto \lambda^{\sum x_i} e^{-n\lambda}$ → "receita bayesiana" multiplicar o núcleo da priori e o da verossimilhança

$\therefore$ $P(\lambda \mid \text{dados}) \propto P(\lambda) \cdot P(\text{dados} \mid \lambda) \Rightarrow P(\lambda \mid \text{dados}) \propto (\lambda^{a-1} e^{-b\lambda}) \times (\lambda^{\sum_1^n x_i} e^{-n\lambda})$

$= P(\lambda \mid \text{dados}) \propto \lambda^{a + \sum x_i - 1} \exp(-(b+n)\lambda) \rightarrow$ similar ao núcleo da Gama!

Partimos da priori Gama e terminamos com um núcleo que tem a forma Gama! (Distribuição a posteriori também é Gama) - Gama é a priori conjugada da Poisson

Portanto Gama $(a + \sum_1^n x_i, b+n)$ é a nossa posteriori ///

Obs: Os parâmetros α' e β' vem da expressão $\int \lambda^{\bar{x} + \alpha' - 1} e^{-(\beta'+1)} d\lambda$ → isso é da letra C!

→ núcleo de uma Gama.

dos 2: integral de fdp sobre o domínio é $= 1$

b) média, moda e variância a posteriori

Priori → média $= \alpha / \beta$
→ moda $= (\alpha - 1)/\beta$, se $\alpha > 1$,
→ var $= \alpha / \beta^2$

posteriori → média $= \dfrac{a + \sum_1^n x_i}{b+n}$
→ moda $= \dfrac{a + \sum_1^n x_i - 1}{b+n}$
→ variância $= \dfrac{a + \sum_1^n x_i - 1}{(b+n)^2}$

c) Idéia → com crença atualizada para λ, como podemos usar para prever uma nova observação $\tilde{x}$? A distribuição a posteriori para essa pergunta é

→ Probabilidade de uma nova observação

$$P(\tilde{x} \mid \text{dados}) = \int P(\tilde{x} \mid \lambda) \times P(\lambda \mid \text{dados})\, d\lambda$$

→ Nossa crença atualizada de $\lambda \Rightarrow \text{Gama}(a + \sum_1^n x_i,\ b+n)$

→ essa mistura gama e Poisson é uma Binomial negativa

→ "A integral que mistura uma verossimilhança de Poiss com priori ou posteriori gama"

Então, como resposta, temos que a distribuição preditiva a posteriori para uma nova observação $\tilde{x}$ é uma Binomial negativa. Com os parâmetros dessa distribuição derivados diretamente da posteriori gama(α', β')

$r = \alpha' = a + \sum_1^n x_i$ Número de falhas (r)

$p = \dfrac{\beta'}{\beta' + 1} = \dfrac{b+n}{b+n+1}$ Probabilidade de sucesso (p)

$$\therefore \tilde{x} \mid \text{dados} \sim NB\left(r = a + \sum_1^n x_i,\ p = \frac{b+n}{b+n+1}\right)$$

Exercício 7) Distribuição a priori : Gama(a, b)

fdp exponencial $p(x_i|\lambda) = \lambda e^{-\lambda x_i}$

verossimilhança $= P(\text{dados}|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \propto \lambda^n e^{-\lambda \sum_i^n x_i}$ → A priori: $p(\lambda) \propto \lambda^{a-1} e^{-b\lambda}$

Encontrar núcleo da posterior:

$P(\lambda|\text{dados}) \propto (\lambda^{a-1} e^{-b\lambda}) \times (\lambda^n e^{-\lambda \sum_i^n x_i})$

simplificando:

$P(\lambda|\text{dados}) \propto (\lambda^{n+a-1} e^{-(b+\sum_i^n x_i)\lambda})$ → núcleo de uma Gama! Gama é a priori conjugada para a verossimilhança da exponencial

$\therefore$ Gamma $(a+n, b+\sum_i^n x_i)$ é a nossa posteriori

b)

Priori (Gama (α, β)) → $\frac{\alpha}{\beta}$ média; → $\frac{(\alpha-1)}{\beta}$, se $\alpha > 1$ moda; → $\frac{\alpha}{\beta^2}$ variância

$\therefore$ Posteriori → $\frac{a+n}{b+\sum_i^n x_i}$ média; → $\frac{a+n-1}{b+\sum_i^n x_i}$ ~~moda~~ moda; → $\frac{a+n}{(b+\sum_i^n x_i)^2}$ variância

c) $P(\tilde{x}|\text{dados}) = \int_0^\infty p(\tilde{x}|\lambda) \cdot P(\lambda|\text{dados})\, d\lambda \Rightarrow P(\tilde{x}|\text{dados}) = \int_0^\infty (\lambda e^{-\lambda \tilde{x}}) \times \left(\frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \lambda^{\alpha'-1} e^{-\beta'\lambda}\right) d\lambda$

na nossa posteriori: $\alpha' = a+n$ e $\beta' = b + \sum_i^n x_i$

Reorganizando a integral

$$P(\tilde{x}|\text{dados}) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \int_0^\infty (\lambda \times \lambda^{\alpha'-1}) \times (e^{-\lambda\tilde{x}} \times e^{-\beta\lambda})\, d\lambda$$

$$= P(\tilde{x}|\text{dados}) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \int_0^\infty \lambda^{\alpha'} e^{-(\beta+\tilde{x})\lambda}\, d\lambda$$

→ como é o núcleo da Gama podemos usar o "truque" propriedade do resultado

$$\int_0^\infty y^{A-1} e^{-By}\, dy = \frac{\Gamma(A)}{B^A}$$

$$\therefore P(\tilde{x}|\text{dados}) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \times \frac{\Gamma(\alpha'+1)}{(\beta'+\tilde{x})^{\alpha'+1}}$$ → $\Gamma(z+1) = z\Gamma(z)$

$$= \frac{\alpha'(\beta')^{\alpha'}}{(\beta'+\tilde{x})^{\alpha'+1}}$$ → FDP distribuição de Lomax ou Pareto tipo 2

$\therefore \tilde{x}|\text{dados} \sim \text{Lomax}(\alpha = a+n, \lambda = b + \sum_i^n x_i)$

— // —

Resolução © Exercício 6 ↓

Exercício 8) Receita Bayes

a)

$$P(\mu \mid \text{dados}) \propto P(\text{dados} \mid \mu) \times p(\mu)$$

A priori $\mu \sim N(\mu_0, \sigma_0^2) \longrightarrow$ núcleo $e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$ de $p(\mu)$

$$P(\text{dados} \mid \mu) \propto \prod_{i=1}^{n} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \Longrightarrow \exp\left(\sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2\right)$$

$$\therefore P(\mu \mid \text{dados}) \propto \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \therefore \mu \sim N(\mu_n, \sigma_n^2)$$

b) Variância posteriori = precisão priori + precisão posterior.

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \therefore \sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

Precisão inverso da variância
Alta precisão baixa variância

Regra Bayesiana: a precisão da crença final é a soma das precisões da crença inicial e da formação vinda dos dados.

média posterior

$$\mu_n = \frac{\left(\frac{1}{\sigma_0^2}\right)\mu_0 + \left(\frac{n}{\sigma^2}\right)\bar{x}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \therefore \mu_n = \sigma_n^2\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right)$$

$\rightarrow$ média ponderada entre a média da crença inicial e a média observada dos dados

c) se $n \to \infty \nearrow \frac{1}{\text{precisão}}$

a variância $\sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$ praticamente vai a 0

analogamente nossa precisão vai para o infinito!

veja $\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \longrightarrow \infty$!

ou seja, a variância da nossa crença sobre μ vai a 0.

Já em $\mu_n = \sigma_n^2\left(\frac{\mu_0}{\sigma_0} + \frac{n\bar{x}}{\sigma^2}\right) \longrightarrow$ converge mais rápido quando $n \to \infty$

Portanto ~~fo~~ a média da posteriori (μ_n) converge para a média amostral ($\bar{x}$)

d) A consequência de $\sigma_0^2 \to \infty$ na variância é

$$\frac{1}{\sigma_n^2} = 0 + \frac{n}{\sigma^2} = \frac{n}{\sigma^2}$$

Ou seja a variância da posteriori σ_n^2 se torna $\frac{\sigma^2}{n}$. A incerteza da crença final vem apenas dos dados

Já para a média

$$\mu_n = \frac{(0)\mu_0 + \left(\frac{n}{\sigma^2}\right)\bar{x}}{0 + \frac{n}{\sigma^2}} = \frac{\left(\frac{n}{\sigma^2}\right)\bar{x}}{\left(\frac{n}{\sigma^2}\right)} = \bar{x}$$

A média da posteriori (μ_n) converge para a média amostral

Exercício 9

a) Caso 1:

Passo 1) verossimilhança

PDF Weibull $= P(x_i \mid k, \lambda) = k \lambda x_i^{k-1} e^{-\lambda x_i^k}$

Verossimilhança $= \prod_{i=1}^{n} k \lambda x^{k-1} e^{-\lambda x^k_i} \xrightarrow[\text{o núcleo (parte que depende de k)}]{\text{simplificando}} P(\text{dados} \mid k) \propto k^n (\prod x_i)^{k-1} e^{-\lambda \sum x_i^k}$

Passo 2) A priori $\to$ gama-inversa (α, β)

Núcleo dessa distribuição é $P(k) \propto k^{-(\alpha+1)} e^{-\beta/k}$

Passo 3) A posteriori

$$P(k \mid \text{dados}) \propto \left(k^n (\prod x_i)^{k-1} e^{-\lambda \sum x_i^k}\right) \times \left(k^{-(\alpha+1)} e^{-\beta/k}\right)$$

$$\Rightarrow P(k \mid \text{dados}) \propto k^{n-\alpha-1} (\prod x_i)^{k-1} e^{-(\lambda \sum x_i^k + \beta/k)}$$

//

Caso 2)

O mesmo processo do caso (1), porém agora com k conhecido e λ parâmetro.
Repetindo o processo temos a seguinte verossimilhança

$$P(\text{dados} \mid k, \lambda) \propto k^n \lambda^n (\prod x_i)^{k-1} e^{-\lambda \sum x_i^k}$$

como k é constante podemos isolar em λ e obter o núcleo da verossimilhança tal que

$$P(\text{dados} \mid \lambda) \propto \lambda^n e^{-\lambda \sum x_i^k}$$

Priori para λ gama inversa (α, β) com núcleo

$$P(\lambda) \propto \lambda^{-(\alpha+1)} e^{-\beta/\lambda}$$

$\therefore$ Nossa posteriori

$$P(\lambda \mid \text{dados}) \propto \left(\lambda^n e^{-\lambda \sum x_i^k}\right) \cdot \left(\lambda^{-(\alpha+1)} e^{-\beta/\lambda}\right)$$

$$\Longrightarrow P(\lambda \mid \text{dados}) \propto \lambda^{n-\alpha-1} e^{-(\lambda \sum x_i^k + \beta/\lambda)}$$

b) e c)

Ambos os casos para (b) e (c) não podem ser resolvidos a mão pois a posteriori não é uma distribuição padrão.

Sendo assim não podemos encontrar a derivada preditiva, média, moda e variância do modo convencional.

Talvez seja possível computacionalmente.